

TP03 - Segmentation des Clients avec KMeans

Contexte

Vous travaillez pour une entreprise de vente au détail et votre mission est de segmenter les clients d'un centre commercial ou d'un magasin en ligne. L'objectif est d'identifier des groupes distincts de clients en fonction de leurs comportements d'achat afin d'optimiser les campagnes de marketing ciblé.

Objectifs de l'exercice

Votre travail consistera à :

1. Préparer les données pour l'analyse.
2. Effectuer une visualisation exploratoire.
3. Implémenter et évaluer l'algorithme de KMeans.
4. Proposer des recommandations basées sur les segments identifiés.

Étapes à suivre

Étape 1 : Préparation des données

- **Importation des données** : Importez le dataset des clients d'un centre commercial depuis Kaggle.
- **Exploration des données** : Étudiez la structure des données, identifiez les variables pertinentes, et traitez les valeurs manquantes si nécessaire.
- **Normalisation des données** : Assurez-vous que toutes les variables sont sur une échelle comparable en appliquant une normalisation ou une standardisation.

Étape 2 : Visualisation exploratoire

- **Analyse exploratoire** : Réalisez une analyse exploratoire des données pour identifier des tendances et des corrélations.
- **Visualisation** : Créez des graphiques pour mieux comprendre la distribution des données et pour anticiper les résultats de clustering.

Étape 3 : Implémentation du KMeans

- **Choix du nombre de clusters** : Utilisez des méthodes comme l'Elbow method, le coefficient de silhouette ou le Gap statistic pour déterminer le nombre optimal de clusters.
- **Application du KMeans** : Implémentez l'algorithme KMeans pour segmenter les clients en différents groupes.
- **Analyse des clusters** : Interprétez les résultats en analysant les caractéristiques principales de chaque cluster.

Étape 4 : Évaluation des clusters

- **Évaluation de la qualité des clusters** : Évaluez la cohésion interne et la séparation entre les clusters obtenus avec l'indice de Silhouette
- **Comparaison avec d'autres méthodes** : Comparez les résultats du KMeans avec ceux obtenus par d'autres algorithmes de clustering comme DBSCAN ou Agglomerative Clustering, et discutez des avantages et inconvénients de chaque méthode.

Étape 5 : Recommandations

- **Proposition de stratégies** : Pour chaque segment de clients identifié, proposez des stratégies de marketing ou d'autres actions spécifiques à mettre en place pour maximiser la satisfaction des clients et augmenter les ventes.

Dataset à utiliser

- [Mall Customer Segmentation Data](#)

Consignes supplémentaires

- **Validation croisée** : Ajoutez un modèle de validation croisée pour évaluer la stabilité des clusters obtenus.
- **Discussion des résultats** : Discutez en profondeur des segments identifiés et de leur pertinence pour l'entreprise.